DOCUMENT RESUME

ED 187 753 TM 800 256

AUTHOR Wise, Steven L.
TITLE An Approach to Sensitive Questions: The Randomized Response Technique.
PUB DATE Apr 80
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Attitude Measures; Behavioral Science Research; College Students; Higher Education; *Questioning Techniques; *Questionnaires; *Research Problems; *Response Style (Tests)
IDENTIFIERS *Randomized Response Technique; Social Desirability Scale (Marlowe Crowne)

ABSTRACT

The randomized response technique, constructed by Warner (1965), was designed to reduce the number of untruthful responses to sensitive questionnaire items. The present study investigated the effectiveness of this technique for groups differing in the tendency to give socially desirable responses to questionnaire items. A questionnaire concerning certain unacademic activities, such as cheating on exams, was given to 158 participants in either a self-report or randomized-response format. The proportion of people endorsing the unacademic activities items was higher using randomized response for all five items, with three items showing significant differences between the two methods. It is suggested that randomized response technique can reduce untruthful responding, but that it should only be used with items known to be sensitive in nature. No evidence was found that social desirability was a factor in participant response to the items. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

- ED187753

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

An Approach to Sensitive Questions:

The Randomized Response Technique

Steven L. Wise

University of Illinois

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
S. Wise

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 800256

Paper presented at the arnual meeting of the American Educational Research Association, Boston, April, 1980.

Abstract

The randomized response technique, constructed by Warner (1965), was designed to reduce the number of untruthful responses to sensitive questionnaire items. The present study investigated the effectiveness of this technique for groups differing in the tendency to give socially desirable responses to questionnaire items. A questionnaire concerning certain unacademic activities, such as cheating on exams was given to 158 participants in either a self-report or randomized-response format. The proportion of people endorsing the unacademic activities items was higher using randomized response for all five items, with three items showing significant differences between the two methods. It is suggested that the randomized response technique can reduce untruthful responding, but that it should only be used with items known to be sensitive in nature. No evidence was found that social desirability was a factor in participant responses to the items.

An Approach to Sensitive Questions: The Randomized Response Technique

When behavioral scientists collect data using questionnaires, they are often faced with the problem of participants giving untruthful responses to some, or all, of the items. This is particularly likely to happen when the items pertain to personal attitudes or behaviors that are perceived by the participant as being sensitive in nature. The tendency to give untruthful responses is a form of "response bias", as described by Hansen, Hurwitz and Madow (1953), and it supposedly is not present when the items are perceived as being relatively innocuous. This response bias results from the reluctance of the participant to reveal personal information about himself.

Warner (1965) has suggested a method designed to reduce or eliminate response bias. Warner's method, which is called the randomized response technique, is based on the assumption that a person is more likely to respond truthfully to a sensitive item if the experimenter or interviewer has no knowledge of the individual's response. In Warner's technique, each item consists of two complementary statements which divide the population into mutually exclusive groups: those that maintain a particular attitude or behavior in question, and those that do not. Thus, each statement requires a dichotomous response (i.e. yes/no or true/false). The items are of the form:

(1) I am a member of group X.

(2) I am not a member of group X.

In Warner's technique, a participant responds to one of the statements of the item according to a random selection known only to him. For instance, the participant may roll a die, with certain sides instructing him to respond to statement 1, and with the remaining sides instructing him to respond to statement 2. The die in this case serves as a randomizing device.

Regardless which statement the participant is instructed to respond to, by endorsing or not endorsing the statement, he is furnishing information about his membership in group X. The experimenter has obtained no direct information about the group membership of an individual participant. However, given a substantial sample size, the probabilities of selecting each statement of the pair, and the proportion of participants who endorse the statement to which they respond, one can estimate the proportion of participants who are members of group X.

Let

π = the true population proportion with membership in group X,

p = the probability of selecting the statement 'I am a member of group X',

n = the number of participants endorsing the selected statement,

N = the total sample size, and

λ = the probability of a 'yes' response to the selected statement.

Then

$$\lambda = \pi\rho + (1-\pi)(1-\rho).$$

By estimating λ as $\hat{\lambda} = n/N$, and solving for π, Warner (1965) derived a maximum likelihood estimate of π, which is an unbiased estimate of the population proportion:

$$\hat{\pi} = \frac{\rho-1}{2\rho-1} + \frac{n}{(2\rho-1)N}, \quad \rho \neq .5.$$

The variance of the estimate is

$$\mathrm{Var}(\hat{\pi}) = \frac{1}{N}\left(\frac{1}{16(\rho-.5)^2} - (\pi-.5)^2\right).$$

Greenberg, Abul-Ela, Simmons, and Horvitz (1969) modified Warner's method by pairing the sensitive statement with an unrelated statement, rather than pairing the sensitive statement with its complement. They suggest that under the Warner method the fact that both statements in the pair pertain to the sensitive topic may arouse suspicion. The unrelated question randomized response technique alleviates this problem.

The variance of $\hat{\pi}$ for the two methods was also investigated. Greenberg et al. (1969), Moors (1971), and Dowling and Shachtman (1975) have shown that in many circumstances the unrelated question method yields a $\hat{\pi}$ with a smaller variance than that obtained with Warner's technique, and thus is more efficient. Furthermore, Greenberg et al. (1969) and Folsom, Greenberg, Horvitz, and Avernathy (1973) have shown that when the population proportion of persons endorsing the

unrelated statement is known, the variance of $\hat{\pi}$ will be less than when the population proportion is unknown.

Let π' be the population proportion of participants for whom the unrelated statement is true. Then

$$\lambda = \rho\pi + \pi'(1-\rho).$$

Again, letting $\hat{\lambda} = n/N$, the maximum likelihood estimate can be obtained, with the value

$$\hat{\pi} = \{\hat{\lambda}-\pi'(1-\rho)\} / \rho .$$

The variance of $\hat{\pi}$ is

$$\mathrm{Var}(\hat{\pi}) = \lambda(1-\lambda) / N\rho^2 .$$

Several studies have used the randomized response technique in the field. Greenberg et al. (1969) used the unrelated question method to estimate the proportion of women giving birth to illegitimate children in North Carolina. The obtained estimate was strikingly similar to that derived from birth certificate data. Goodstadt and Gruson (1975) compared the unrelated question method with a traditional self-report measure to estimate drug use in a sample of high school students. The randomized response technique gave a significantly higher estimate than the self-report method. The authors suggest that traditional questionnaires underestimate incidence of drug use. In estimating the prevalence of induced abortion in Taiwan, Rider, Harper, Chow, and Chi I-Cheng (1976) showed that the randomized response technique gave a significantly higher $\hat{\pi}$ than that found in a personal interview. Wiseman, Moriarity, and Schafer (1976) found significant differences between the randomized response technique and a self-report instrument with statements of a political,

moral, or racial-prejudicial nature. However, they failed to find differences between personal interviews and the randomized response technique. They suggested that a lack of confidence in the randomizing device on the part of a substantial proportion of the participants may have partially accounted for the nonsignificant findings. Zdep and Rhodes (1976) compared randomized response and self-report formats in a survey of child abuse among a large sample of parents. They found a substantially higher estimate using the randomized response technique. Locander, Sudman, and Bradburn (1976) compared four interview methods (face-to-face, telephone, self-report, and randomized response) and found that the randomized response technique yielded the least amount of underreporting of socially undesirable acts, such as being charged with drunken driving. However, the randomized response technique did not outperform the other interview methods in terms of the overreporting of socially desirable behaviors.

The nature of the difference between estimated proportions obtained by different data gathering techniques has been discussed by Wiseman (1972). He also noted that the techniques offering the most response privacy to the participant tend to obtain the highest proportion of socially undesirable responses. Under the randomized response technique, one would expect participants to feel relatively more free to respond to an item in a truthful way, regardless of how socially undesirable their response may seem. In this manner, suggests Wiseman, the randomized response technique may be used to help alleviate the problem created by social desirability response biases.

Assuming that there are individual differences in the tendency to respond according to social desirability, one might expect that the randomized response technique would eliminate the greatest amount of response bias among participants with a high tendency to be biased toward giving the socially desirable response. That is, individuals with a high tendency should show the greatest difference between $\hat{\pi}$ obtained by a self-report measure and $\hat{\pi}$ obtained by use of the randomized response technique. This difference between the two estimates would serve as an indicator of bias reduction by the randomized response technique.

The social desirability tendency should have a positive correlation with the amount of bias reduction for an individual. However, since direct data are not available for individual respondents under the randomized response technique, groups of respondents differing in social desirability tendency would have to be compared in order to demonstrate the relationship between social desirability and the amount of bias reduction.

The purpose of this study was to compare a group of participants showing a high social desirability tendency with a group showing a low tendency, in order to demonstrate that the randomized response technique produces more bias reduction in the group with the high social desirability tendency. Bias reduction was evaluated by comparing the proportion of participants endorsing sensitive items in a self-report format, with the estimated proportion obtained under the unrelated question randomized response technique. An individual's

social desirability tendency was measured by the Marlowe-Crowne Social Desirability Scale.

Three types of results were expected. First, for all participants combined, $\hat{\pi}$ for the randomized response technique was expected to be significantly higher than the $\hat{\pi}$ obtained with the self-report. This prediction followed directly from the assumption that participants were more likely to respond truthfully under the randomized response technique condition and therefore the proportion estimated to have the undesirable characteristic was greater. Second, for both response techniques combined, $\hat{\pi}$ was expected to be higher for the group with the low social desirability tendencies. This prediction followed from the assumption that truthful responding was more likely in the low tendency group. Finally, there was a significant interaction expected between social desirability and response technique, with the difference in $\hat{\pi}$ between the response techniques being greater in the high tendency group than in the low tendency group, since it was in the high tendency group that the bias toward untruthful responding was likely to be greater.

Method

Participants

The participants were 158 introductory psychology students taking part in the study as partial fulfillment of their course requirements. Warner (1965) has shown that when the sample sizes are equal, the standard error of $\hat{\pi}$ obtained with the randomized response technique will be larger than that obtained with a conventional self-report technique. Participants were therefore assigned to either the randomized

response or the self-report condition, with 60% of the sample serving in the randomized response technique condition, and the rest serving in the self-report condition. Since increasing the sample size in the randomized response technique results in a decrease in the standard error of $\hat{\pi}$, using different numbers of participants in the two response conditions would result in standard errors of $\hat{\pi}$ that are more comparable.

Procedure

The participants were recruited for a study investigating unacademic behaviors and attitudes in students from a small eastern university. They served in groups, ranging in size from one to ten. All of the members of a given group served in the same response condition. The participants were seated at least 1.5 m apart, to minimize the possibility of participants observing the responses of other group members. The experimenter stayed in the room during the entire session.

Two instruments were administered to the groups, in counterbalanced order between groups. The participants' social desirability tendencies were measured by the Marlowe-Crowne Social Desirability Scale. Standard administration procedures were used for this scale. The second instrument, constructed for this study, was called the Survey of Unacademic Activities. This instrument consisted of five statements assumed to be "sensitive", describing behaviors and attitudes which students may not readily attribute to themselves, such as cheating on exams or stealing materials from libraries. Three of

the items dealt with specific behaviors, while the other two were attitudinal in nature. The instrument used a true-false format and the items were all keyed such that item endorsement represented the socially undesirable response.

The Survey of Unacademic Activities for the randomized response condition is shown in Appendix A. Each sensitive statement was paired with a related statement dealing with participants' telephone numbers. The randomizing devices used were decks of twenty index cards, consisting of 14 cards with "Answer Statement A" printed on one side, and six cards with "Answer Statement B" printed on them. The participants were instructed, for each pair of statements, to shuffle the deck, randomly draw a card, and follow the instructions on the card. Thus, the probability of a participant responding to the sensitive statement was .70. For the self-report condition, participants were asked to fill out a version of the Survey of Unacademic Activities which contained only the five sensitive statements.

After both instruments had been administered in the randomized response technique condition; the participants were asked to rate how confident they felt that their responses were unknown to the experimenter. They used a four-point rating scale ranging from "not at all confident" to "completely confident". This participant rating is similar to that used by Wiseman et al. (1976) in which no differences were found between the randomized response technique and self-report method, due partially to the fact that many of the participants did not have confidence in the randomizing device.

Statistical Analysis

The data for each statement were analyzed in a two-factor design, with social desirability as one factor, and response technique as the other. Since there were five sensitive statements, five separate analyses were performed. The social desirability factor had two levels, low and high, with participants within each response technique being assigned using a median split on the distribution of scores on the Marlowe-Crowne Social Desirability Scale. The response technique factor had the two levels of self-report and randomized response technique.

For each of the four conditions, an estimated proportion of participants endorsing the sensitive statement, $\hat{\pi}$, was computed. Appendix B describes the calculations and formulae used in the statistical analysis. Three hypotheses were tested.

(1) H_0: $(\pi_{rl} + \pi_{rh}) - (\pi_{sl} + \pi_{sh}) \geq 0$.

(2) H_0: $(\pi_{rl} + \pi_{sl}) - (\pi_{rh} + \pi_{sh}) \geq 0$.

(3) H_0: $(\pi_{rh} - \pi_{sh}) - (\pi_{rl} - \pi_{sl}) \geq 0$.

The first hypothesis tested whether there was any difference between $\hat{\pi}$ obtained under the two response techniques. The second hypothesis tested for differences in $\hat{\pi}$ between the low and high social desirability tendency groups. The third hypothesis tested for the interaction of social desirability and response technique.

Results

Procedural Factors

The median split for the distribution of Marlowe-Crowne scores placed participants with a score greater than 15 in the high social desirability group, and the remaining participants in the low social desirability group. This resulted in slightly unequal group sizes within each response condition. The self-report condition had 36 participants in the high group and 27 in the low group, while the randomized response technique condition had 46 and 49 in the high and low groups, respectively.

Since the number of participants choosing to serve in each group was uncontrolled, groups were nonsystematically assigned to both response condition and instrument administration order. Since strict counterbalancing was therefore not maintained, several checks were made for any factors that may have systematically affected the results. Order of instrument presentation was not found to be related to response condition, r (156) = -.02, p>.05. Marlowe-Crowne scores for the two response conditions were not found to be significantly different, t (146) = -1.56, p>.05. Although average group size was found to be significantly different for the two response conditions, t (156) = 4.05, p<.001, group size was not related to Marlowe-Crowne score, r (156) = -.09, p>.05.

The average confidence rating for the participants in the randomized-response condition was 2.56, with 48 of the participants (50.5%) stating that they were either "not at all confident" or

"slightly confident" that the experimenter did not know which statement they had answered. The remaining 47 participants (49.5%) gave a rating of either "very confident" or "completely confident". These ratings were found to be unrelated to either group size, r (93) = .00, $p>.05$, or instrument presentation order, r (93) = .13, $p>.05$.

Differences Among Groups

The estimated proportion of participants ($\hat{\pi}$) endorsing the sensitive item, broken down by experimental group for each of the items in the Survey of Unacademic Activities is shown in Table 1. The overall pooled estimates vary considerably, with a high $\hat{\pi}$ of .31 on Item 4 (marking in library books) and a low of .09 on Item 2 (removing library materials illegally).

When social desirability groups were combined for each response method, the randomized response technique yielded a greater $\hat{\pi}$ than with the self-report technique, for all five items. The largest differences occurred on Item 1 (cheating on a college exam) and on Item 5 (some cheating is justifiable).

Table 2 summarizes the statistical analyses for the items. In testing for differences among the four experimental groups, only Item 2, U_0' (3) = 9.64, $p<.05$, and Item 5, U_0' (3) = 10.79, $p<.05$ showed significant differences.

The results of the three planned comparisons among the experimental groups are also shown in Table 2. In comparing response methods, significant differences were found for Item 1, z = 2.47, $p<.01$, Item

2, $z = 2.71$, $p<.01$, and Item 5, $z = 2.54$, $p<.01$. In comparing social desirability groups, no significant difference was found for any item. There was also no significant interaction between response method and social desirability.

Discussion

It was predicted that $\hat{\pi}$ for the randomized response technique would be larger than the $\hat{\pi}$ obtained for the self-report method. The differences between $\hat{\pi}$'s for the response methods were in the expected direction for all five items. Significant differences were found for items 1, 2, and 5. The differences were not significant for items 3 and 4, although item 3 approached significance ($p<.10$).

Sensitivity of items and sample size are both major influences on the significance of the difference between response methods. If sensitivity is defined as the tendency of an item to elicit untruthful responding, items which are not sensitive in nature would not be expected to show a difference in $\hat{\pi}$ between response methods, since the amount of untruthful response bias removed by the randomized response technique should be minimal. It was expected that the items in this study would be sufficiently sensitive as to result in differences between response methods. The finding that items 3 and 4 did not show significant differences between response methods suggests that they might not have been as sensitive for this population as items 1, 2, and 5. Also, a relative lack of confidence in the randomized response technique by half of the sample may have

contributed to the lack of response method effect for the less sensitive items 3 and 4.

Another possible explanation for the nonsignificant method differences is that the standard error of $\hat{\pi}$ is partly a function of sample size. Previous studies finding significant differences between response methods (Goodstadt & Gruson, 1975; Wiseman et al., 1976) used sample sizes much larger than those used in this study. Since the standard error of $\hat{\pi}$ obtained is larger than with a self-report measure, and the method is more awkward and time consuming to use, the randomized response technique should probably be used only for items which are known in advance to be sensitive in nature.

The expected difference in $\hat{\pi}$ between social desirability groups was not found, suggesting that social desirability was not a factor in participant responses to the Survey of Unacademic Activities. This result is surprising, since the items of this scale were chosen to contain a socially desirable response, as well as being chosen to be sensitive in nature. It was assumed that the participants would see item endorsement as the socially undesirable response.

It was also expected that there would be found an interaction between social desirability and response method. None of the items showed this interaction. This finding does not support the proposal by Wiseman (1972) that the response bias removed by the randomized response technique is related to social desirability.

The assumption that participants would see item endorsement as the socially undesirable response may have been unfounded. Cheating

on exams, removing library materials, and trying to get ahead academically any way that one can may be positively valued by some in the college-student culture. If this is the case, then the socially desirable response may not be the same for all participants. This could lead to a lack of difference between social desirability groups, since two participants in the same social desirability group, but with different personal values, could respond to the academic activities statements in opposite ways. Although it is not expected that all of the members of a group hold similar values, the socially desirable response must be seen as such by a large majority of group members, otherwise, the term "socially desirable response" becomes ambiguous. Such ambiguity may have been present in this study.

However, the absence of social desirability effects may also be explained by possible invalidity of the Marlowe-Crowne Social Desirability Scale. Bradburn and Sudman (1979) describe a study of the scale which concludes that individual differences in Marlowe-Crowne scores are related to actual differences in individual activities, experiences, and attitudes, rather than differences in social desirability response tendency. If this is the case, then groups differing in social desirability response tendency may not actually have been formed in this study.

Conclusions

The randomized response technique is useful in obtaining more accurate estimates of the prevalence of behaviors and attitudes which

respondents see as being sensitive in nature. Although it is necessarily more time consuming to employ, and needs more respondents to produce reasonably stable estimates, use of randomized response is well worth the effort in many instances. However, if the items are not seen as sensitive by the respondents then other, more economical, methods of data collection should be used.

When using the randomized response technique one should be careful to explain to the respondents how the method works. Higher confidence in the respondents of the anonymity of their responses will probably result in more truthful responses and consequently more accurate estimates.

Reduction in response bias due to use of the randomized response technique is not related to Marlowe-Crowne scores. Although there may be some question as to whether or not the Marlowe-Crowne measures social desirability tendency, whatever is measured is not related to the use of randomized response.

References

Bradburn, N. M., & Sudman, S. Improving interview method and questionnaire design. San Francisco: Jossey-Bass, 1979.

Dowling, T. A., & Shachtman, R. H. On the relative efficiency of randomized response models. Journal of the American Statistical Association, 1975, 70, 84-87.

Edwards, A. L. The social desirability variable in personality assessment and research. New York: Dryden Press, 1957.

Folsom, R. E., Greenberg, B. G., Horvitz, D. G., & Abernathy, J. R. The two alternate questions randomized response model for human surveys. Journal of the American Statistical Association, 1973, 68, 525-530.

Goodstadt, M. S., & Gruson, V. The randomized response technique: a test on drug use. Journal of the American Statistical Association, 1975, 70, 814-818.

Greenberg, B. G., Abul-Ela, A. A., Simmons, W. R., & Horvitz, D. G. The unrelated question randomized response model theoretical framework. Journal of the American Statistical Association, 1969, 64, 520-539.

Hansen, M. H., Hurwitz, W. N., & Madow, W. G. Sample survey methods and theory, volume II. New York: John Wiley and Sons, Inc., 1953.

Locander, W., Sudman, S., & Bradburn, N. An investigation of interview method, threat, and response distortion. Journal of the American Statistical Association, 1976, 71, 269-275.

Marascuilo, L. A. Large sample multiple comparisons. Psychological Bulletin, 1966, 65(5), 280-290.

Moors, J. J. Optimization of the unrelated question randomized response model. Journal of the American Statistical Association, 1971, 66, 627-629.

Rider, R. V., Harper, P. A., Chow, L. P., & I-Cheng, C. A comparison of four methods for determining prevalence of induced abortion, Taiwan. American Journal of Epidemiology, 1976, 103(1), 37-50.

Warner, S. L. Randomized response: a survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 1965, 60, 63-69.

Wiseman, F. Methodological bias in public opinion surveys. Public Opinion Quarterly, 1972, 36, 105-109.

Wiseman, F., Moriarity, M., & Schafer, M. Estimating public opinion with the randomized response model. Public Opinion Quarterly, 1976, 39, 507-513.

Zdep, S. M., & Rhodes, I. N. Making the Randomized response technique work. Public Opinion Quarterly, 1976, 40, 531-537.

Table 1

Estimated Proportion of Item Endorsement ($\hat{\pi}$) for the Five Sensitive Items in the Survey of Unacademic Activities for Different Conditions

Item	Conditions						Overall Pooled Estimate
	Self-Report			Randomized Response			
	High S.D.	Low S.D.	Combined Groups	High S.D.	Low S.D.	Combined Groups	
(1) Cheated on a College Exam	.22	.15	.19	.39	.41	.40	.25
(2) Removed Library Materials Illegally	.08	.04	.06	.17	.34	.25	.09
(3) Will Cheat if Necessary to Get Good Grades	.19	.25	.22	.39	.30	.34	.21
(4) Marked in Library Books	.22	.37	.29	.29	.44	.37	.31
(5) Cheating on Exams is Justifiable	.28	.11	.21	.39	.44	.41	.25

Table 2

Statistics for the Overall Effect of Groups and Planned Comparisons for Five Items

Item	Effect			
	Overall Effect (U_0')	Planned Comparisons (z)		
		Response Method	Social Desirability	Interaction Method and Social Desirability
(1) Cheated on a College Exam	6.85	2.47**	- .29	- .575
(2) Removed Library Materials Illegally	9.64*	2.71**	.87	-1.51
(3) Will Cheat if Necessary to Get Good Grades	4.49	1.41	- .16	.90
(4) Marked in Library Books	3.74	.78	1.62	.01
(5) Cheating on Exams is Justifiable	10.79*	2.54**	- .65	-1.28

*$p<.05$.

**$p<.01$.

Appendix A

A Survey of Unacademic Activities

For each of the pairs of items, answer one statement according to the instructions given to you. Indicate in the space provided whether the statements are true or false as they pertain to you personally.

A. I have cheated on a college exam on at least one occasion.

_____ B. My telephone number ends in either a 1,2,3, or 4.

A. I have removed library materials "illegally" on at least one occasion.

_____ B. My telephone number ends in either a 6,7,8, or 9.

A. In order to get ahead after college, I will cheat if necessary to get good grades.

_____ B. My telephone number ends in either a 3,4,5, or 6.

A. I have marked in books belonging to a college library on at least one occasion.

_____ B. My telephone number ends in either a 1,3,5, or 9.

A. I feel that cheating on exams is justifiable in some instances.

_____ B. My telephone number ends in either a 0,4,6, or 8.

Appendix B

Statistical Analysis

The experimental design produced four independent estimates of $\hat{\pi}$, and an estimate of the variance of each $\hat{\pi}$:

$$\hat{\pi} = \hat{\lambda}_{sl} = n_{sl} / N_{sl} \qquad \text{var}(\hat{\pi}_{sl}) = \hat{\pi}_{sl}\,(1-\hat{\pi}_{sl})/N_{sl}$$

$$\hat{\pi} = \hat{\lambda}_{sh} = n_{sh} / N_{sh} \qquad \text{var}(\hat{\pi}_{sh}) = \hat{\pi}_{sh}\,(1-\hat{\pi}_{sh})/N_{sh}$$

$$\hat{\pi} = \frac{\hat{\lambda}_{rl} - \pi'(1-\rho)}{\rho} \qquad \text{var}(\hat{\pi}_{rl}) = \frac{\hat{\lambda}_{rl}(1-\hat{\lambda}_{rl})}{N_{rl}\rho^2}$$

$$\hat{\pi} = \frac{\hat{\lambda}_{rh} - \pi'(1-\rho)}{\rho} \qquad \text{var}(\hat{\pi}_{rh}) = \frac{\hat{\lambda}_{rh}(1-\hat{\lambda}_{rh})}{N_{rh}\rho^2}$$

Marascuilo (1966) has shown that if a weighted average of a set of independent estimates is taken, using the reciprocal of $\text{var}(\hat{\pi}_i)$ as the weight $\hat{W}_i$, the resulting pooled estimate of π, $\hat{\pi}_0$, may be used to form a statistic $\underline{U}_0'$ which has a χ^2 distribution, with k-1 degrees of freedom.

$$\hat{\pi}_0 = \sum_{i=1}^{k} \frac{\hat{\pi}_i}{\text{var}(\hat{\pi}_i)} \Big/ \sum_{i=1}^{k} \frac{1}{\text{var}(\hat{\pi}_i)} = \sum_{i=1}^{k} \hat{W}_i \hat{\pi}_i \Big/ \sum_{i=1}^{k} \hat{W}_i .$$

$$U_0' = \sum_{i=1}^{k} \frac{(\hat{\pi}_i - \hat{\pi}_0)^2}{\text{var}(\hat{\pi}_i)} = \sum_{i=1}^{k} \hat{W}_i (\hat{\pi}_i - \hat{\pi}_0)^2 .$$

This statistic will be used to test the null hypothesis H_0: $\pi_{sl} = \pi_{sh} = \pi_{rl} = \pi_{rh} = \pi$. In addition, comparisons among the parameters may be tested for significance. A comparison is a linear combination of parameters which is estimated from the linear combination of sample statistics:

$$\hat{\psi} = \sum_{i=1}^{k} a_i \hat{\pi}_i .$$

The estimated variance of $\hat{\psi}$ is given by:

$$\text{var}(\hat{\psi}) = \sum_{i=1}^{k} a_i^2 \text{var}(\hat{\pi}_i) .$$

Marascuilo (1966) has presented a multiple-comparison procedure analogous to the Scheffé procedure. In Marascuilo's procedure, the null hypothesis H_0: $\psi = 0$ is rejected if the interval

$$\hat{\psi} \pm \sqrt{\chi^2_{k-1}(1-\alpha)} \quad \sqrt{\text{var}(\psi)}$$

does not include 0. Equivalently, H_0 is rejected if $\hat{\psi}^2/\text{var}(\hat{\psi})$ exceeds the critical value of $\chi^2_{k-1}(1-\alpha)$.

This procedure, like the Scheffé procedure, is conservative with respect to the probability of a Type I error, and permits all possible comparisons to be tested with a fixed experimentwise Type I error rate of α. However, when planned orthogonal comparisons are to be tested, as in this study, more powerful tests can be performed if we note that $\hat{\psi}/\sqrt{\text{var}(\hat{\psi})}$ has a unit normal distribution under H_0. Thus, an appropriate critical value of z may be chosen for each comparison.

The three comparisons of major interest are given below:

(1) H_0: $(\pi_{rl} + \pi_{rh}) - (\pi_{sl} + \pi_{sh}) \geq 0$ (main effect of methods)

$$\hat{\psi}_1 = \hat{\pi}_{rl} + \hat{\pi}_{rh} - \hat{\pi}_{sl} - \hat{\pi}_{sh}$$

(2) H_0: $(\pi_{rl} + \pi_{sl}) - (\pi_{rh} + \pi_{sh}) \geq 0$ (main effect of social desirability)

$$\hat{\psi}_2 = \hat{\pi}_{rl} + \hat{\pi}_{sl} - \hat{\pi}_{rh} - \hat{\pi}_{sh}$$

(3) H_0: $(\pi_{rh} - \pi_{sh}) - (\pi_{rl} - \pi_{sl}) \geq 0$ (interaction of methods and social desirability)

$$\hat{\psi}_3 = \hat{\pi}_{rh} - \hat{\pi}_{sh} - \hat{\pi}_{rl} + \hat{\pi}_{sl}$$

For each of the comparisons listed above, the variance estimate is

$$\text{var}(\hat{\psi}) = \text{var}(\hat{\pi}_{rh}) + \text{var}(\hat{\pi}_{sh}) + \text{var}(\hat{\pi}_{rl}) + \text{var}(\hat{\pi}_{sl}).$$